Whitney Alderman & Gazal Arora

Assignment 0: Tokenizer
A program in C that takes String inputs of any size and separates it into specific tokens.

Types of Tokens Addressed in this program:
• Decimal Integer: a digit character followed by any number of digit characters
• Octal Integer: '0' character followed by any number of octal digits (0-7)
• Hexadecimal Integer: "0x" (or "0X") followed by any number of hexadecimal digits (0-9, af, A-F)
• Word: consists of an alphabetic character followed by any number of alphanumeric characters
• Floating Point: a decimal integer that has a '.' at any position other than the last. It may optionally contain an exponent in scientific notation at the end
• C Operators: 44 operators that get identified as operator token

We have also implemented all 3 extra credits to include the following tokens:
• C Keywords: The 32 C Keywords like int, enum, float, etc are identified as C keywords from the input string
• Single Quote: ex: ‘this is a single quote token’
• Double Quote: ex: “this is a double quote token”

Comments in C: These are skipped as they are comments. Comments starting with // end when they hit a \n or \0. Comments starting with /* end when they hit */

Compilation: Compile as follows in the command line
$ gcc -o tokenizer tokenizer.c

Run: Must run with two arguments in command line
./tokenizer "input string"